Journal of Web Librarianship, 7:465-476, 2013
ISSN: 1932-2909 print / 1932-2917 online
DOI: 10.1080/19322909.2013.825148

Data-Visualization and Rapid Analytics: Applying Tableau Desktop to Support Library Decision-Making

Sarah Anne Murphy©

Abstract

Data visualization offers librarians the ability to better manage, explore and present information collected by various individuals throughout a library organization. This article discusses The Ohio State University Libraries experiments with Tableau, a sophisticated data-visualization and rapid analytics software. Tableau allows librarians to blend and leverage data collected from a number of disparate sources, including transaction logs, Google Analytics, and e-resource usage reports. The article provides context for incorporating data-visualization into the OSU Libraries assessment program and shares examples of visualizations created for two data-analysis projects. The benefits of blending and simultaneously viewing visualizations of data from multiple sources are articulated and explored. The article concludes with a short discussion of potential future projects for visualizing library data using Tableau Desktop.

Keywords

data visualization; data analysis; library assessment; Google Analytics; transaction logs; learning management systems; dashboards; Tableau software

Traditional research requires practitioners to define a question or problem first, but what if you don't know what your questions are until you actually see the data, or what questions might be possible? Further, what if data analysis answered the original question and generated many more? Rapid analytic and data-visualization software offers librarians the flexibility to explore and make-sense of large quantities of data. No expert knowledge of computer programming or tools such as Google Charts or Microsoft Excel is required. As libraries seek to capitalize on the opportunities provided by big data, such software offers librarians the ability to efficiently and effectively make sense of large data sets with limited human and financial resources. This article will explore The Ohio State University Libraries experiments with Tableau Desktop, a rapid-fire business intelligence software, in effort to better manage and present data collected by various individuals across a number of library departments. It will begin by defining data-visualization, emphasizing its benefits, and investigating its current application within the academic library community. It will continue by providing context for the incorporation of data-visualization into the OSU libraries' assessment program. Visualizations for two data-analysis projects will then be shared, showcasing the benefits of blending multiple data sources to effectively make better

© Sarah Anne Murphy, Coordinator of Assessment, The Ohio State University Libraries, 255C Thompson Library, 1858 Neil Avenue Mall, Columbus, OH 43210. murphy.465@osu.edu

sense of large data sets. Last, the article will conclude with a short discussion of potential future projects for visualizing library data using Tableau Desktop.

Data is inherently messy, especially as libraries change platforms, upgrade systems, and archive diverse datasets over long periods of time. To analyze long-term trends, legacy data must be located and then carefully cleaned to allow accurate comparisons or calculations. Librarians must know who is collecting and storing what data, to avoid costly duplication of effort. As staff retire or move on to new opportunities, libraries also risk losing valuable tacit knowledge concerning how a dataset was structured and why, especially if employees failed to keep detailed documentation. Librarian and staff comfort and skills with tools such as Microsoft Excel spreadsheets and Microsoft Access databases vary significantly across most institutions. Further, the skills required to understand and analyze data differ significantly from the skills necessary to present data in a creative, compelling way.

"Data can often be sterile," notes Nathan Yau, " but only if we present it that way." (Yau 2009, 1.6) Data-visualization makes "stories visible and bring[s] them to life." (Few 2009a, 5) The term encompasses both information and scientific visualization, showcasing "all types of visual representations that support the exploration, examination, and communication of data." (Few 2009a, 12) Visual representations of data are much more powerful than rote presentation of numbers. While tables require us to read and consider the meaning and relationship of each individual value presented, visualizations allow us to process many values concurrently. Further, humans can more efficiently and effectively process a graph than a table of numerical text. Thus, visualizations allow librarians to recognize trends, spot patterns, and identify exceptions.

Literature Review

While the concept of data-visualization is not new, researchers have been presenting abstract data visually for more than 200 years, technology now supports real-time visualization of large data sets. Academic libraries are starting to recognize the value of visualization tools to support agile decision-making. For example, using visual design principles, along with PERL scripts, a MySQL database, and Google Chart Tools, Emily Morton-Owens and Karen Hanson created a management data dashboard for the New York University Health Sciences Libraries (NYUHSL). (Morton-Owens and Hanson 2012) The dashboard drew together disparate datasets, from EZproxy logs, Google Analytics, Libraryh3lp, and the libraries' interlibrary loan system. It notably revealed what e-resources were being used, segmented by department and other user groups. While a proof-of-concept project, Morton-Owens and Hanson demonstrated that collective metrics presented together on one dashboard may indeed help NYUHSL librarians to learn more about their user populations, celebrate success, and realize areas requiring improvement.

To visualize a device-lending programs and make better data-driven management decisions, Joyce Chapman and David Woodbury with the North Carolina State University Libraries' experimented with Protovis, an open-source Java-Script library and Microsoft Excel. (Chapman and Woodbury 2012) They plotted variables such as wait time for circulating iPads, Kindles, projectors, camcorders, and other devices, along with capacity lending hours for various laptop models over the 2010-2011 academic year. Their visualizations quickly illustrated devices with average wait times longer than one week, informing decisions to purchase or discontinue select devices. They also illustrated demand fluctuations for laptops

throughout the day and semester, informing circulation staffing decisions and lending policy revisions by device type.

Chapman with Cory Lown also illustrated the use of data-visualization software tools in an article discussing efforts to improve data analysis projects at the North Carolina State University Libraries. (Chapman and Lown 2010) To generate real-time, web-based reports for the libraries' custom federated search utility, custom PHP scrips were written and then fed into Google's Chart Tools . The resulting visualizations and reports provided valuable information on usage patterns and influenced redesign of the search results page. Other projects which created visualizations to aid analysis included a custom transaction log analysis of NCSU's Library Course Tools application, and an analysis of longitudinal trends in patron use of reference services.

Data -Visualization at The Ohio State University Libraries

The Ohio State University Libraries formally established an assessment program with the appointment of an assessment coordinator in February 2012. While working to enhance a portal that provides a central location for librarians and staff to access data about library operations and services, the coordinator researched best-practices for centrally storing and communicating data in real-time. She realized that many divisions and units within the Libraries might benefit from visual analytics and began to research software or services such as IBM's Many Eyes or Google's Chart Tools. (IBM Research 2013; Google 2013) The coordinator recommended that the libraries pilot Tableau Desktop, a breakthrough, visual analytics software which enables users to rapidly explore, analyze, and present abstract data.

Tableau differs significantly from other sophisticated data-visualization software and programs by integrating querying, exploration, and visualization of data into a single process. The software defaults ensure that users create effective visual presentations of data, using color, form, spatial position, and motion. (Few 2009b, 35-49) Further, after selecting relevant dimensions or attributes and related measures or values, Tableau automatically recommends visualizations which would most effectively represent the data. Features such as its date algorithm, enable users to adjust time-series by year, quarter, month, or day, with one or two clicks, saving time and enhancing their ability interact with data. Tableau supports two qualities required for effective information visualization: the "ability to clearly and accurately represent information and [users] ability to interact with it to figure out what the information means." (Few 2009a, 55) Tableau enables users the flexibility to respond to questions as they arise, eliminating the need for pre-defined questions or programming assistance from information technology staff.

With Tableau, users can add variables, re-sort, aggregate values, highlight values or filter data while creating a visualization. Users can also shift effortlessly among alternative visualizations without needing to reformat or process their data. Tableau offers users the ability to blend disparate data sources regardless of whether the data was stored in a Microsoft Excel spreadsheet or Access database. With an upgrade from Tableau Desktop Personal to the Tableau Desktop Professional edition, data may be extracted and manipulated from multiple sources, including, Oracle, MicrosoftSQL Server, MySQL, PostgreSQL, Google Analytics, or any ODBC Version 3.0 compliant system. (Tableau 2013) This offers the potential for real-time, or close to real-time analysis.

By blending data from multiple sources, academic libraries can leverage data collected from transaction logs, Google Analytics, e-resource usage reports, and other sources to create meaningful dashboards within Tableau. Such dashboards may provide real-time management information that informs library operations, web design, marketing activities, or instruction. Two examples of efforts to visualize data to support decision-making for all aspects of the library's virtual presence follow. The first presents three years of transactional log data highlighting usage of library guides embedded in the university's learning management system (LMS). The second provides a simple snapshot of Google Analytics data concerning usage of the libraries' web site by language. This information is presented in concert with international student enrollment figures and the number of undergraduates with a declared non-English language or area studies major.

Carmen Library Link: Visualizing Uptake of Library Guides Embedded in the University's Learning Management System

Since 2007, The Ohio State University Libraries has used its locally established Carmen Library Link system to automatically deliver 178 library guides to more than 5,000 courses in Carmen, the university's LMS. (Black 2008; Black and Blankenship 2010; Murphy and Black) To better understand trends in student uptake of library guides embedded within the university's LMS, the transaction logs for the 2009 to 2012 academic years were pulled. These logs provide usage information for each library guide and include the name and number of the course to which the guide was assigned, the library guide name and ID, the date and time for each instance the guide was used, and the email address for the librarian responsible for maintaining the guide. Any information which might personally identify a student was not queried from the system.

To provide a more meaningful overview, the data was blended with profiles for librarians and academic departments. It was then further segmented by course-level, librarian, and the subject teams to which librarians are assigned. The first dashboard provides an overview of overall guide use by month, a crosstab table displaying guide usage by the academic department in which the course resided, and a graph showing uptake of library guides created by two librarians on the science subject team. (Figure 1) Note that each librarian is represented by his or her own color on the line graphs. Any individual accessing this dashboard can also add or remove data from the 'Trend By Librarian' graph by using the filters on the right side of the screen.

[PLACE FIGURE 1 APPROXIMATELY HERE]

First glance at the 'Overall Guide Use By Month' section of the dashboard reveals a significant outlier with the potential to distort the interpretation and analysis of the data. Tableau saves time by allowing users to simply hover over any data point to automatically display the values represented by that data point. (Figure 2) Hovering over the major outlier on the 'Overall Guide Use by Month' graph reveals that in September 2011, the library guides created by one librarian were used 32,636 times. Follow-up revealed that the librarian was the instructor of record for a large enrollment introductory level science course, and required students to use her guides to complete assignments. This fact is important when considering the quantitative story of guide uptake, especially since it visually appears that use of other librarian's guides is not significant when the outlier value is included. Figure 3, for instance, illustrates how the "Overall Guide Use by Month' and 'Trend By Librarian' graphs change significantly when the outlier is excluded.

[PLACE FIGURES 2 AND 3 APPROXIMATELY HERE]

To further focus attention where the libraries might redesign or improve marketing of library guides to select or strategic departments, the 'Overall Guide Use By Department' table provides the number of times library guides were used overall by each academic department. Interested users can sort or re-sort each column in this table. Figure 4 provides a second dashboard with both a trend chart and heat maps to help librarians and administrators visualize changes in the uptake of library guides by year, and reflect on the activities which may have influenced guide use. The heat maps use size and color to help illustrate trends in guide usage by department. Larger light bulbs, for example, indicate that a guide was used more in one particular year, in relation to another. Guides with higher aggregate use for a given year are greener in the 'Number of Visits by Academic Department' table. A global filter was applied to this visualization. This means changes to the checkbox filters on the right side of the screen are reflected on all three graphs. The highlight feature was also used to help better understand the uptake of library guides for one academic department in relation to the uptake of guides by all academic departments serviced by these two librarians.

[PLACE FIGURE 4 APPROXIMATELY HERE]

Tableau rapidly processed more than 258,000 transaction logs to create the textured visualizations shown in Figures 1-4. The next example showcases the power of Tableau to blend smaller, seemingly unrelated data sets, presenting a more holistic picture of an audience for a library's web site.

Basic Analytics: Visualizing Usage of a Library Web Site in Languages Other Than English

Nearly 5,000 international students attend The Ohio State University, representing 9.8% of the entire graduate and undergraduate student body. The university also supports a number of respected Title VI language and area studies programs, including the Department of East Asian Languages and Literatures and the Center for Slavic and East European Studies. To better understand the languages spoken or studied by individuals using the libraries' homepage, a basic language report for the years 2010 to 2012 was exported from Google Analytics, with the understanding that American English would likely be the default browser setting of users of the OSU Libraries website. In isolation, this Google Analytics report might not be very valuable. In concert with graphs summarizing data from two other sources, the Libraries may obtain a better understanding of the use of its website by populations with unique needs.

Figure 5 provides a dashboard summarizing data from the three sources listed above. The map presents data from the University Registrar's 15th Day Enrollment report. This map uses color to help users visualize total enrollment by country. Note that China appears in the next shade of green to indicate that Chinese students represent the second largest student population at The Ohio State University. Hovering over or clicking on China opens a tooltip that shows in fall of 2012, OSU enrolled 3,423 graduate and undergraduate students from China. (Figure 6) The horizontal bar chart illustrates the language report pulled from Google Analytics. After English, which is excluded to allow other language browser settings to be visualized, Chinese is the second most popular browser language default, followed by Korean and then Spanish. When the English is added back into the 'Language Report from Google Analytics' graph, and the bar chart is re-sorted in descending order by 'Average Visit Duration,' it is interesting to note that individuals using Spanish language settings in their browser visit more pages, and spend more time on the Libraries' web site. (Figure 7) The 'Language and Area Studies Majors at The

Ohio State University' table completes the dashboard visualization of languages spoken or studied by users of the libraries website. This data was gathered via a custom query of the university's student information system to determine undergraduate and graduate enrollment in language and area studies majors. The table shows that after International Studies, Spanish is the language and area studies program with the second highest enrollment at OSU. Thus, the insights from simultaneously viewing and exploring these three seemingly unrelated datasets may help to inform librarian's future content and design decisions for the library web site.

[PLACE FIGURES 5, 6 AND 7 APPROXIMATELY HERE]

Discussion

Tableau software offers a practical solution to a common academic library problem: how to efficiently explore and make sense of large sets of data with limited human and financial resources. The ability to seamlessly interact with data by quickly adding or excluding variables, segmenting, sorting, highlighting, and other actions significantly enhances our ability to make sense of large volumes of quantitative information. Tableau's drag-and-drop functionality also makes it deceptively easy to shift from one information visualization to another. Built-in tools aid the thought process, freeing users from wasting time while toggling back and forth with another software program to identify outliers, or remove irrelevant lines from a data set.

The Ohio State University Libraries' assessment program is only just beginning to experiment with and explore Tableau's many layered features. The dashboards created for the two examples above confirmed that Tableau software may be used to make library data meaningful and accessible to library staff across departments. Like Adobe, Tableau software has a freely available application that allows individuals to open and view visualizations created in Tableau Desktop, without having to own a copy of the production software itself. The ability to filter data remains functional in Tableau Reader, and individuals may export raw data from the visualizations to an Access database.

The assessment program, however, has not experimented with perhaps one of the most powerful features of Tableau: the ability to harvest and visualize data in real-time from a large database such as a library catalog or interlibrary loan system. (Tableau 2013) With usage data for print and electronic resources coming from a number of different systems, future projects might explore using Tableau to blend, filter, segment, and make sense of large sets of collection use data in real-time. Creative visualizations of library collections information may serve as a valuable communication tool, especially for librarians engaging with faculty to develop collections or administrators tasked with justifying funding. With installation of Tableau's Server product, such visualizations might be embedded on library web pages or shared on an iPAD during a faculty or graduate student consultation.

Tableau might also be employed to gather and analyze data illustrating an academic library's contribution to student success. With ACRL's recent *Value of Academic* Libraries initiative and corresponding *Assessment in Action: Academic Libraries and Student Success* learning community, the ability to blend and visualize relevant library and student datasets may be of interest to a number of academic libraries. (ACRL 2013) Future visualizations might examine uptake of library guides and overall student GPA in courses with a writing component for which a librarian provided at least one library instruction session.

Finally, while Tableau is relatively simple to use, individuals interested in creating more layered, complex visualizations with calculated fields and other functions may need to invest a considerable amount of time in learning and mastering the software. Few libraries are actively using the product. Further, creativity is required to translate Tableau's training materials which are almost exclusively geared for the business environment into library lexicon. For instance, usage may be substituted for revenue and profit might be reconsidered as a measure of cost per use. It is useful to seek a local Tableau Users Group consisting of individuals from the business, non-profit, and local government community who are actively applying Tableau to visualize and understand their large data sets. The Ohio State University Libraries assessment program, however, has determined that the functionality afforded by Tableau, however, is worth the learning investment and will continue to use the software to support its assessment activities.

References

Association of College and Research Libraries. 2013. "ACRL Value of Academic Libraries." http://www.acrl.ala.org/value/ (accessed January 18, 2013).

Black, Elizabeth L. 2008. "Toolkit Approach to Integrating Library Resources Into theLearning Management System." *Journal of Academic Librarianship* 34 (6): 496-501.

Black, Elizabeth L. and Betsy Blankenship. 2010. "Linking Students to Library Resources through the Learning Management System." *Journal of Library Administration* 50: 458-467.

Chapman, Joyce and Cory Lown. 2010. "Practical Ways to Promote and Support Collaborative Data Analysis Projects." *cod4lib* 12: 1-10.

Chapman, Joyce and David Woodbury. 2012. "Leveraging Quantitative Data to Improve a Device-Lending Program." *Library Hi Tech* 30 (2): 210-234.

Few, Stephen. 2009a. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Oakland, CA: Analytics Press. p. 5.

Few, Stephen. 2009b. "Thinking With Our Eyes." In *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, Stephen Few, 35-49. Oakland, CA: Analytics Press.

Google. 2013. "Google Chart Tools." https://developers.google.com/chart/ (accessed January 18, 2013).

IBM Research. 2013. "Many Eyes." http://www-958.ibm.com/software/analytics/manyeyes/ (accessed January 18, 2013).

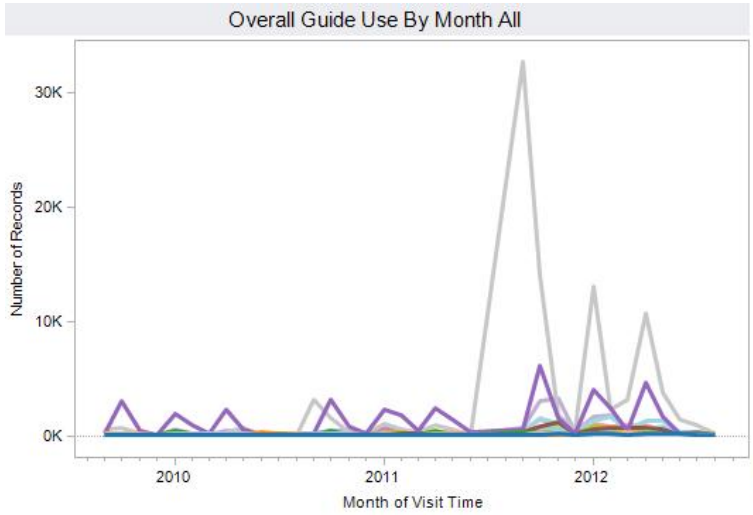
Morton-Owens, Emily and Karen L. Hanson. 2012. "Trends at a Glance: A Management Dashboard of Library Statistics." *Information Technology and Libraries* 31 (3): 36-51.

Murphy, Sarah Anne and Elizabeth L. Black. 2013. "Embedding Guides Where Students Learn: Do Design Choices and Librarian Behavior Make a Difference?" *Portal: Libraries and the Academy* In review.

Tableau Software. 2013. "Technical Specifications – Tableau Desktop." http://www.tableausoftware.com/products/desktop/specs (accessed January 18, 2013).

Yau, Nathan. 2009. "Seeing Your Life in Data." In *Beautiful Data*, eds. Toby Segaran and Jeff Hammerbacher, 1.6. Sebastopol, Calif.; Farnham: O'Reilly.

Figure 1. Dashboard With Overall Guide Use by Month and Academic Department.



Overall Guide Use By Academic Department

	Visit Time				
Academic Department	2009	2010	2011	2012	Grand Tot..
Chemistry	1,418	6,383	50,199	32,309	90,309
English	6,233	13,696	24,256	17,992	62,177
Education: Teaching and Le..	681	1,456	3,415	5,293	10,845
Women's, Gender, and Sexu..	44	163	2,492	2,007	4,706
Biology	96	598	1,748	2,004	4,446
Comparative Studies	35	271	1,604	2,459	4,369
Engineering	97	1,032	972	2,197	4,298
Pharmacy	163	843	1,317	1,918	4,241
Nursing	264	1,280	1,338	1,121	4,003
Sociology	88	436	1,039	1,549	3,112
Mathematics	116	697	1,198	973	2,984
History	301	1,130	886	627	2,944
Social Work	115	505	1,059	1,171	2,850
Psychology	407	929	947	509	2,792

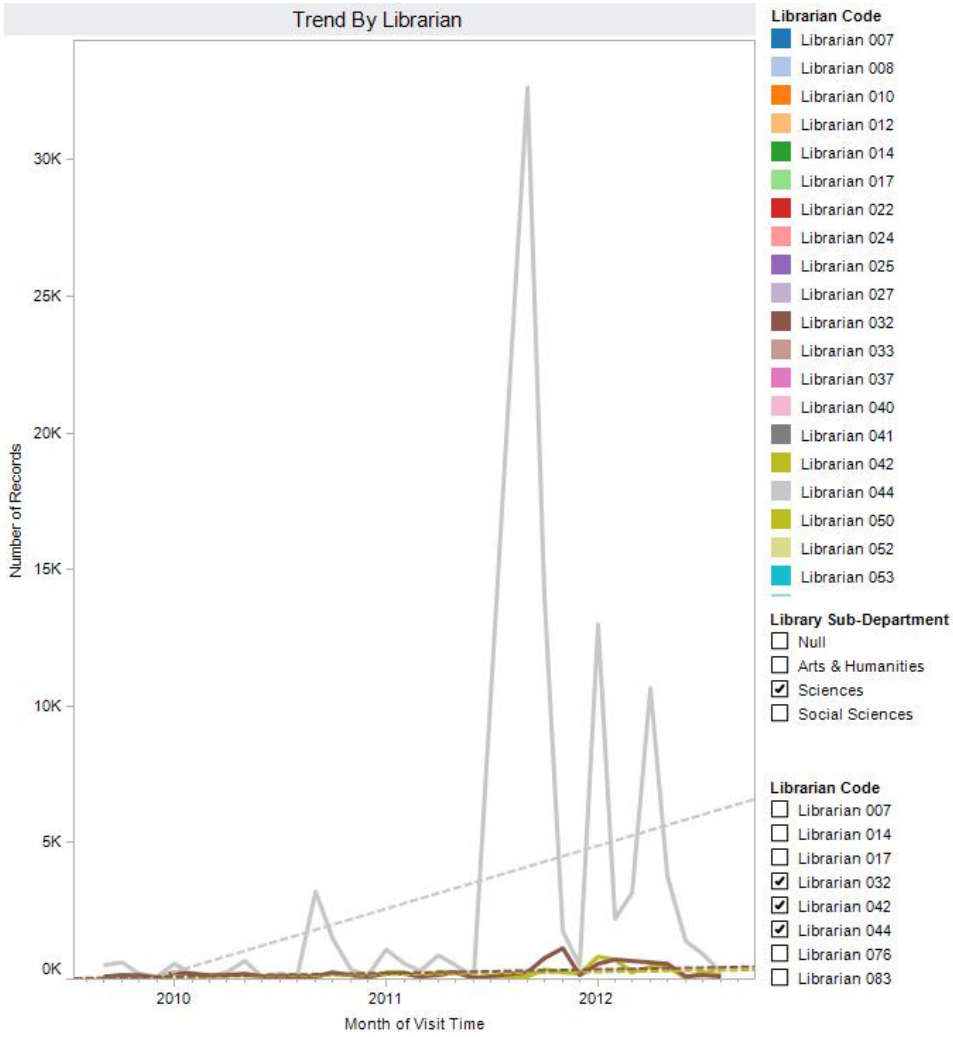


Figure 2. Pop-up Window With Values for Outlier.

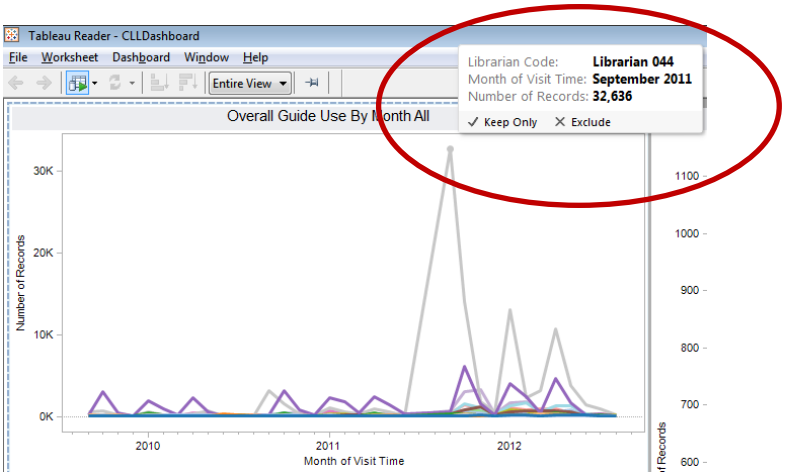
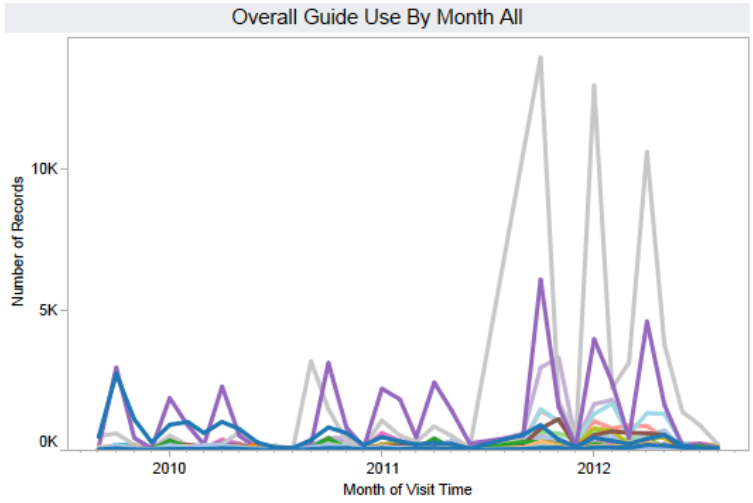


Figure 3. Dashboard With Outlier Value Excluded



Overall Guide Use By Academic Department

	Year of Visit Time				
Academic Department	2009	2010	2011	2012	Grand Total
Chemistry	1,418	6,383	50,199	32,309	90,309
English	6,233	13,696	24,256	17,992	62,177
Education: Teaching and Learn..	681	1,456	3,415	5,293	10,845
Women's, Gender, and Sexuali..	44	163	2,492	2,007	4,706
Biology	96	598	1,748	2,004	4,446
Comparative Studies	35	271	1,604	2,459	4,369
Engineering	97	1,032	972	2,197	4,298
Pharmacy	163	843	1,317	1,918	4,241
Nursing	264	1,280	1,338	1,121	4,003
Sociology	88	436	1,039	1,549	3,112
Mathematics	116	697	1,198	973	2,984
History	301	1,130	886	627	2,944
Social Work	115	505	1,059	1,171	2,850
Psychology	407	929	947	509	2,792

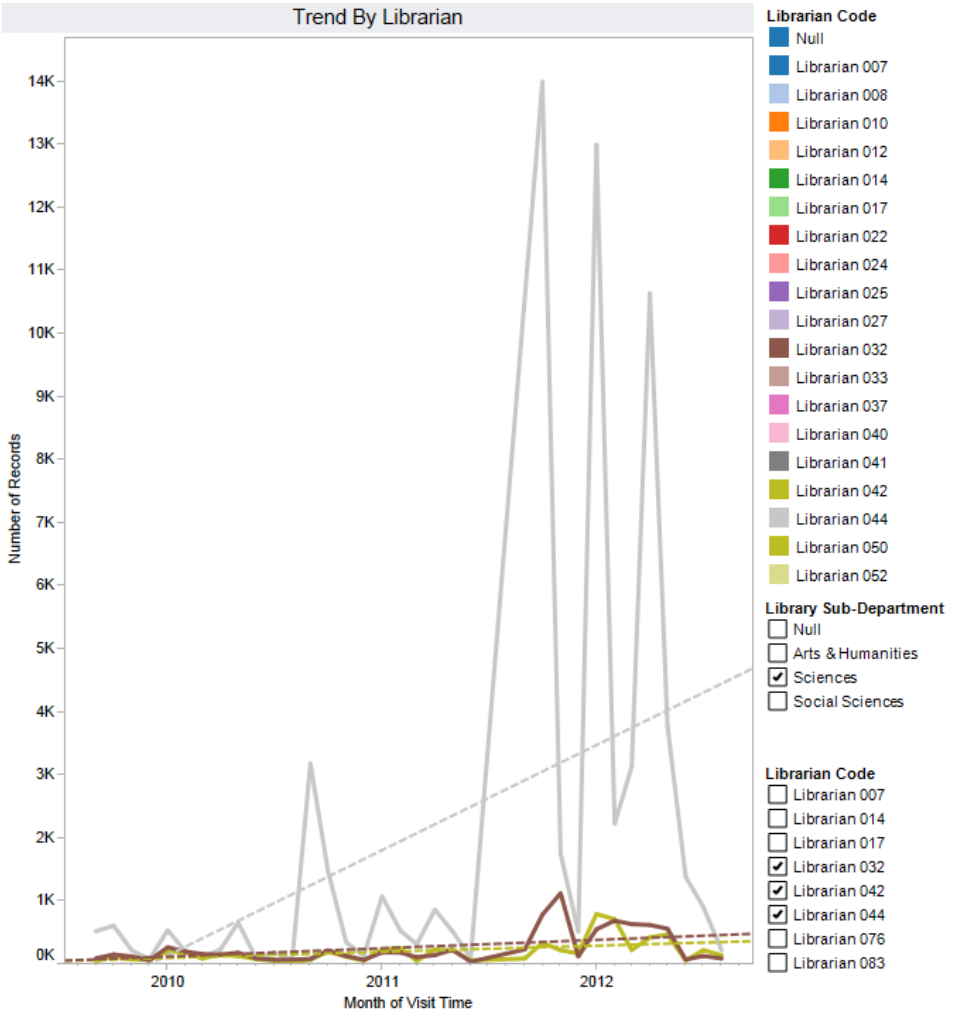


Figure 4. Alternative Visual by Librarian and Academic Department

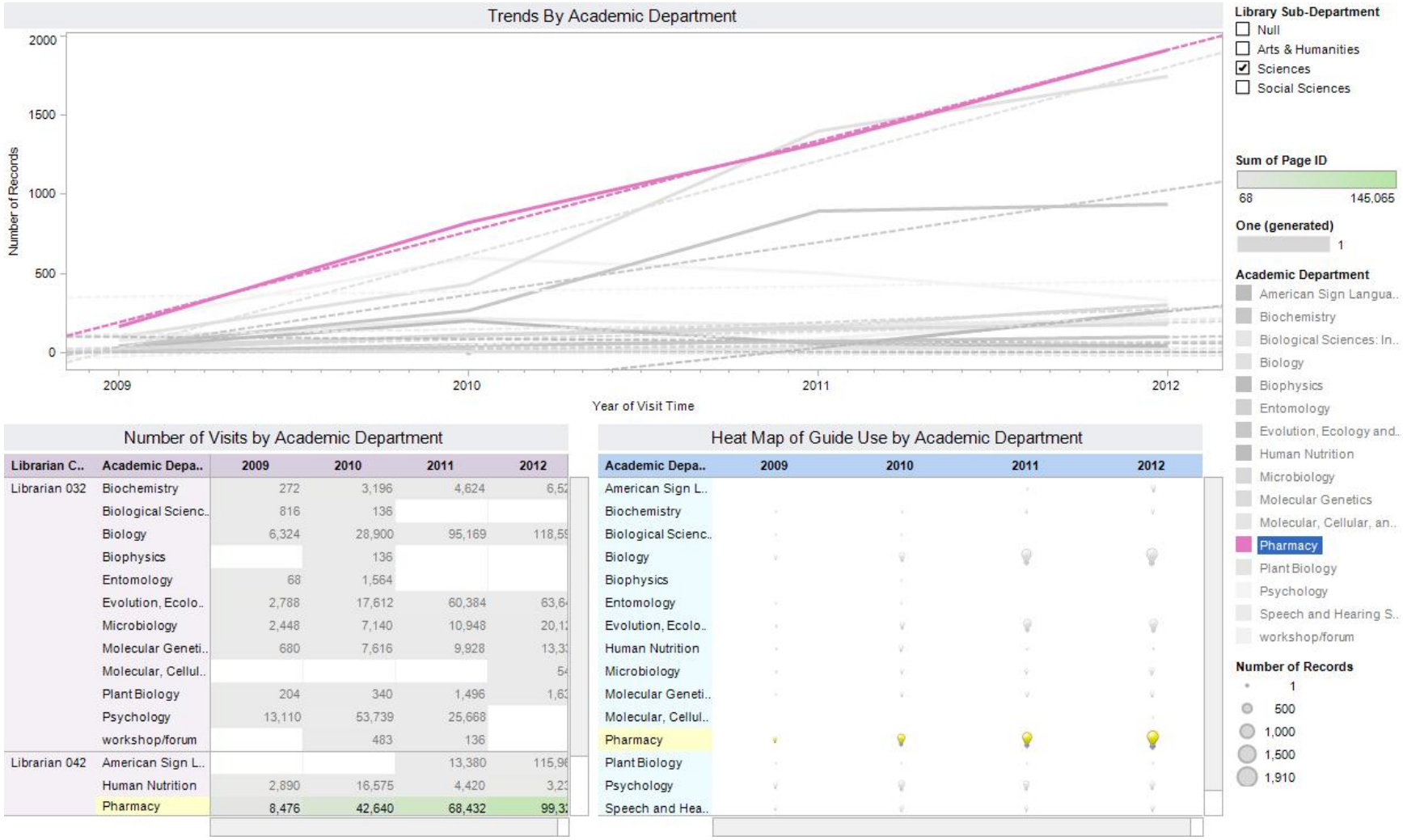


Figure 5. Dashboard of Google Analytics Language Report, With Visuals from the OSU 15th Day Enrollment Report, and Data Regarding Language and Area Studies Majors at The Ohio State University

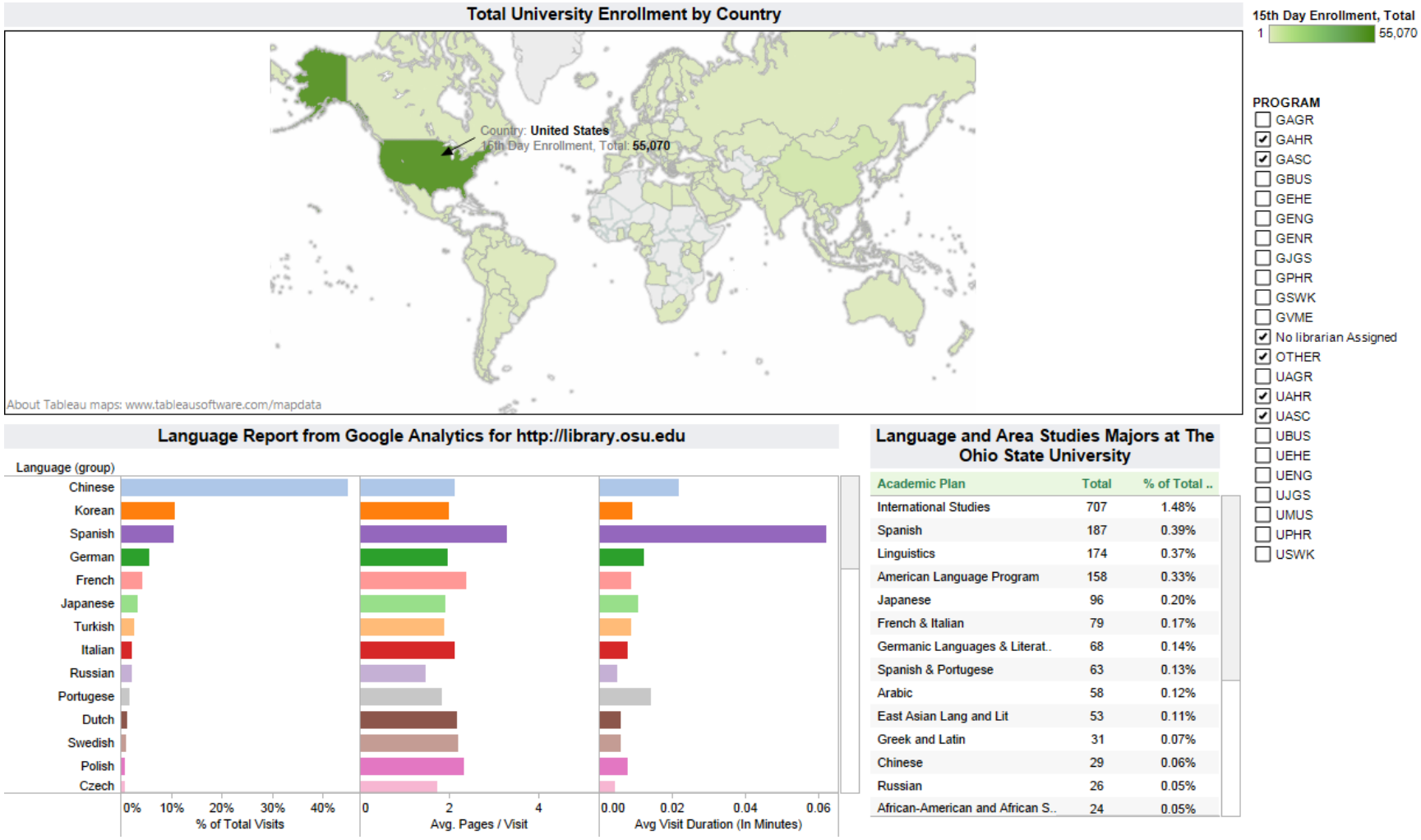


Figure 6. Tooltip Showing Additional Information for China.

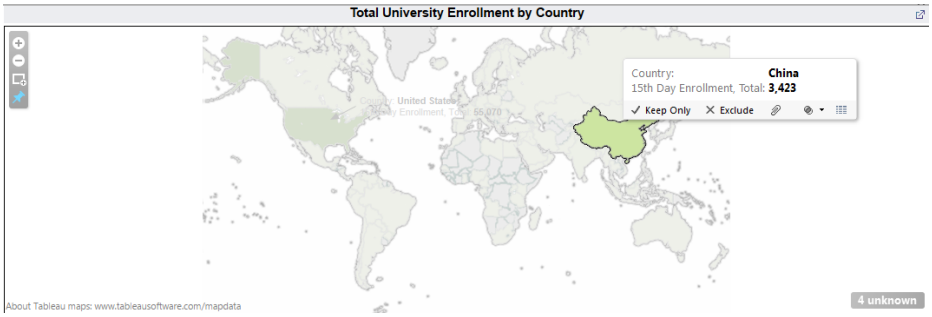


Figure 7. Language Report from Google Analytics Resorted by Average Visit Duration in Minutes